ADAPTIVE SILOUETTE EXTRACTION AND HUMAN TRACKING IN COMPLEX AND DYNAMIC ENVIRONMENTS

Xi Chen, Zhihai He, Derek Anderson, James Keller, and Marjorie Skubic

Department of Electrical and Computer Engineering
University of Missouri, Columbia, MO, 65203, USA

ABSTRACT

Extracting a human silhouette from an image is the enabling step for many high-level vision processing tasks, such as human tracking and activity analysis. Although there are a number of silhouette extraction and human tracking algorithms proposed in the literature, most approaches work efficiently only in constrained environments where the background is relatively simple and static. In this work, we propose to address the challenges in silhouette extraction and human tracking in a real-world unconstrained environment where the background is complex and dynamic. We extract features from image regions, accumulate the feature information over time, fuse the high-level knowledge with low-level features, and build a time-varying background model. We develop a fuzzy decision process to detach foreground moving objects from the human body. Our experimental results demonstrate that the algorithm is very efficient and robust.

Index Terms **-** Image segmentation, silouette extraction, human tracking, fuzzy logic.

1. INTRODUCTION

Silhouette extraction, namely, segmenting the human body from the background, is often the first step and also the enabling step for many high-level vision analysis tasks, such as human tracking and activity analysis. One central task in silhouette extraction is background modeling. Once a background model is established, those image regions with significantly different characteristics from the background are considered as foreground objects. In [1], an LMeds (Least Median of Squares) method is used to construct the background model. The differencing function proposed in [2] is used to extract moving human silhouettes. In [3], a non-parametric kernel density estimation method is employed to model and subtract the background. The adaptive background updating in that approach is able to deal with small motion in the background scene, such as moving tree branches. In [4], a silhouette extraction approach is developed in a YUV color space that is able to detect shadows and segment the human body from the background.

Although there are a number of algorithms developed in the literature for silhouette extraction and human tracking, most of them have targeted and work efficiently only in controlled environments, where the lighting condition is good, and the background is simple and nearly static. In this work, as a part of our project to develop automated video-based activity analysis for eldercare, we propose an efficient and robust silhouette extraction and human tracking algorithm in a real-world, unconstrained environment. More specifically, we will use video cameras to collect information about elderly residents' daily activity at home. From these videos, we will extract important activity information to perform automated functional assessment and detect abnormal events, such as people falling on the floor. To design an efficient video activity monitoring system for eldercare, there are two major issues we need to consider. The first one is privacy protection. Our approach is to extract the human body from the background and fill the corresponding image region with white pixels so as to block the identifying features. The second issue is activity analysis. Our approach is to extract feature information from the human silhouette and develop statistical models, such as Hidden Markov Models, to model and recognize activities. In this paper, we address the silhouette extraction problem.

1.1. Major Challenges

In real-world environments, such as in-home eldercare monitoring, an efficient and robust silhouette extraction and human tracking algorithm has to deal with the following challenges: (1) The lighting condition is time-varying. The change in lighting could be caused by changing sunlight, passing vehicles outside the apartment, indoor lights turned on or off, adjustment of blinds, etc. (2) The background is dynamic. Objects in the background, such as furnitures and appliances are being moved by residents. (3) Silhouette extraction should be accurate and robust. This is because the subsequent vision analysis tasks, such as activity modeling and abnormal event detection, are totally based on the feature information extracted from the silhouette. This implies that the extracted silhouette should be closely conform to the human body. For example, a simple foreground segmentation algorithm may include a chair moved by the person as part of the silhouette. This will create a potential problem because the activity analysis module will detect a sudden change in the silhouette, from a vertical slim human body to horizontally stretched silhouette (human + chair), and may report this change as an abnormal event. According to our study, this is one of the major sources of false alarms in video-based activity monitoring.

Corresponding Author: Zhihai He, HeZhi@missouri.edu.

1.2. The Proposed Approach

We formulate the silhouette extraction as an adaptive classification problem. High-level knowledge is fused with low-level feature-based classification to handle a time-varying background, and a fuzzy decision process is used to detach the moving objects from the human silhouette. Our experimental results demonstrate that the proposed silhouette extraction and human tracking scheme is works efficiently and robustly.

The rest of the paper is organized as follows. Section 2 gives an overview of the whole system. Section 3 describes feature extraction and foreground-background segmentation. In Section 4, we develop an algorithm for adaptive background update. The fuzzy decision process for detaching moving objects is presented in Section 5. Section 6 presents the expermental results. Section 7 concludes the paper.

2. OVERVIEW OF THE ALGORITHM

We assume that the camera is fixed. We formulate silhouette extraction as an adaptive classification problem. The input video frame in partitioned into small blocks, for example, 4×4 blocks. For each block, a classification decision is made: the block belongs to the human silhouette or not. To this end, we extract invariant features from the image blocks, as illustrated in Fig. 1. Based on these features, we build a statistical model for the background and classify the image blocks into two categories: foreground and background. Because the background is time-varying, the background model and the classifier should be adaptive. However, background adaptive is also risky. For example, if a person sits still or sleeps in the couch for a long time, say hours, the adaptive background model will consider the person as part of the background. This is not acceptable because of unprotected privacy. To solve this problem, we fuse the high-level knowledge obtained from human tracking with the low-level feature-based classification so as to guide the background update. At this stage, the foreground may still contain objects other than the human silhouette, such as objects moved by or attached to the person. To address this issue, we develop a fuzzy decision process to detach these objects from the human silhouette. In the following sections, we explain the proposed silhouette extraction scheme in detail.

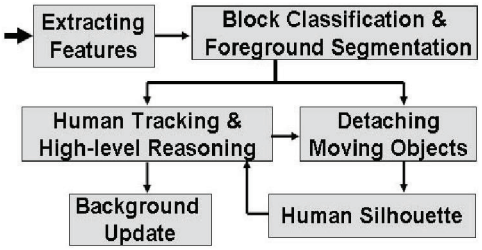

Figure 1: Block diagram of the proposed system for silhouette extraction and human tracking.

3. FEATURE EXTRACTION

We extract features in the RGB color space. Two feature variables: chromaticity and brightness distortions, are used to classify the foreground and background [5]. The color model used here separates the brightness from the chromaticity components as shown in Fig. 2. Here, E_i is the color vector for pixel i in the RGB space. I_i is the observed color vector. Brightness distortion a_i is defined as follows:

$$a_i = arg \min_{a_i} ||I_i - a_i E_i||^2. \quad (1)$$

The color distortion is defined by

$$CD_i = ||I_i - a_i E_i||. \quad (2)$$

The brightness distortion is very helpful in detecting shadows. It has been observed that shadows cause brightness distortion but little color distortion. The foreground and background classification is based on the following observation: image blocks in the background should have little change in their color distortion. In the proposed adaptive background modeling and classification scheme, we use the image data in the past Δ frames to compute the joint distribution of these two feature variables so as to build a background model. For each block in the current frame, we compute the feature vector $[a_i, CD_i]$. Based on this feature vector and the background model, we classify the image block into two categories: foreground or background. Fig. 3 shows the results of foreground and background segmentation using this feature extraction and classification scheme.

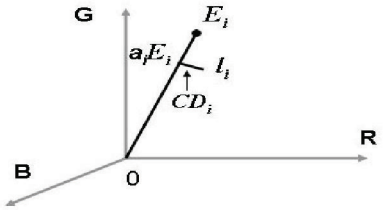

Figure 2: Feature variables in the RGB color space.

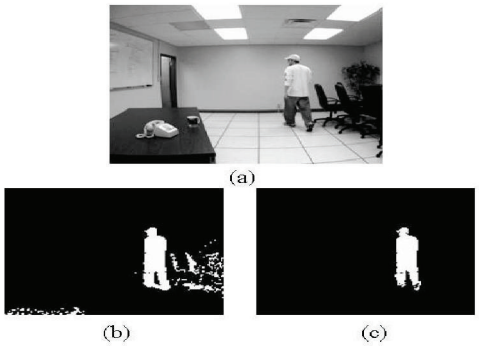

Figure 3: (a) Original image, (b) Silhouette without shadow suppression and (c) Result with shadow suppression.

4. KNOWLEDGE-BASED ADAPTIVE BACKGROUND UPDATE

It should be noted that the value of Δ, i.e., the size of the shifting window, should be appropriately chosen. If Δ is too small, the background is updated very fast; the human body, or parts of the human body, could be easily updated as background when the person remains still longer than Δ frames. In a residential environment, a person could remain still for minutes or even hours. Therefore, the time duration of the shifting window should also be quite large. However, if the time duration Δ is too large, the background is updated very slowly. If a new object is introduced into the background or an object is moved, before the background is updated, this object will be classified into the foreground. Therefore, the foreground silhouette will contain both the human body and the object. As explained in Section 1.1, this will create a significant problem and potentially generate a false alarm. To solve this problem, in this section, we propose to utilize the high-level knowledge about the human motion to guide the adaptive update of the background model.

Our basic idea is to track the human body and predict its region in the scene. The image blocks which contain the human body should be updated very slowly such that the human body would not be updated as background. Those blocks outside the body region can be updated much faster to make sure that new objects are quickly absorbed into the background. Fig. 4 plots the centroid of the human body in frame n, denoted by (x_n, y_n), when the person moves around in the room. Let (w_n, h_n) be the width and height of the bounding box of the human body in frame n. Define $\mathbf{V_n} = (\mathbf{x_n}, \mathbf{y_n}, \mathbf{w_n}, \mathbf{h_n})$. We use the following linear prediction model

$$\mathbf{V_{n+1}} = \sum_{\mathbf{i=0}}^{\mathbf{L}} \mathbf{a_i V_{n-i}}, \tag{3}$$

to predict $\mathbf{V_{n+1}}$ for the human body in frame $n+1$. $\mathbf{V_{n+1}}$ then defines a bounding box, which should contain the human body if the prediction is accurate. For those blocks inside, we set the value of Δ to be much larger than the remaining blocks.

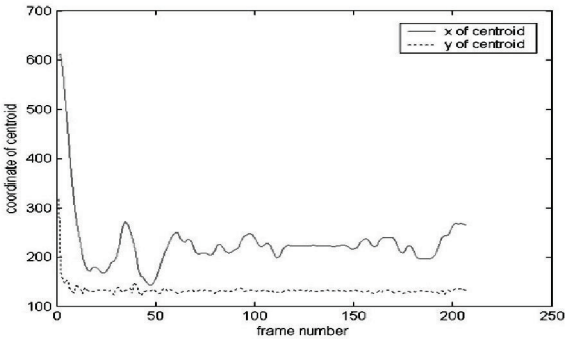

Figure 4: Centroids of the human body.

5. DETACHING THE MOVING OBJECTS USING FUZZY LOGIC

This module is designed to detach objects being moved by the person from the human silhouette. Detaching the objects from the human silhouette is a challenging task because of the lack of automated reasoning about what is human being and what is not. Certainly, sophisticated human recognition and identification algorithms could be used. However, these algorithms are often computationally intensive and not robust in a dynamic environment like homes. Therefore, in this work, we propose to a simple yet efficient algorithm for object detachment. Our scheme is based on a fuzzy decision system, which fuses multiple sources of information together for decision making.

Suppose we are working frame n and the silhouette in frame $n-1$ has been correctly extracted, denoted by O_{n-1}. Let the foreground image region be O_n, which might contain moving objects. Our fuzzy inference is based on the following observations: (1) If an image block in O_n belongs to the human body, it should have a high possibility of finding a good match in O_{n-1}. We use the SAD (sum of absolute difference) to measure the "goodness" of matching. (2) If a lot of blocks in its neighborhood have good matches in O_{n-1}, it is a high possibility that this block also belongs to O_{n-1}, the human body. (3) If this block is far from the predicted position of the human body centroid, as described in Section 4, it is a low possibility that this block belongs to the human body.

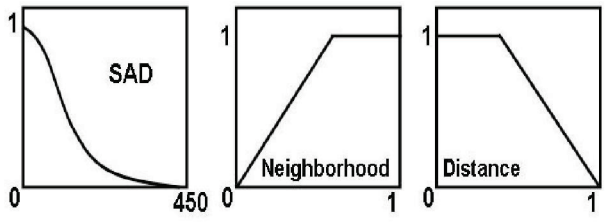

Figure 5: Membership functions of feature variables for the fuzzy inference.

Based on these observations, we extract the following feature variables for each block in O_{n-1}: (1) SAD in motion matching; (2) the fraction of neighboring blocks having a good match in the human body; (3) the distance between this block and the predicted human body centroid. We collected distributions of these three feature variables from test video clips and built membership functions for these fuzzy variables, as illustrated in Fig. 5. The fuzzy decision engine could fuse this information to evaluate how likely the coming block belongs to the human silhouette. A fuzzy rule can be set as follows: *IF SAD is small AND the fraction of neighboring silhouette blocks beloning to the human body is large and the distance from the centroid is small, THEN the new block is more likely to be a silhouette block.*

5.1. Experimental Results

We have implemented the proposed system for human silhouette extraction and evaluated its performance in a sim-

Table 1: Numerical performance of silhouette extraction.

True silhouette	92.1%
True background	99.7%
False silhouette	0.3%
False background	7.9%

ulated residential environment. Fig. 6 shows the human silhouette extraction with and without human movement tracking and adaptive background update. In this experiment, the person walks in and sits still on the chair for several minutes. The algorithm is still able to keep the person in the silhouette while updating background very fast, so that the bag on the floor is classified into the background.

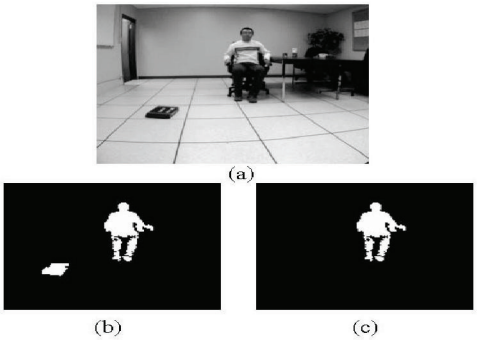

Figure 6: Background-foreground segmentation with (c) and without (b) human movement prediction and background update.

Fig. 7 shows the extraction results with and without moving objects detachment using the fuzzy inference system. We can see the proposed silhouette extraction scheme works very effectively in various dynamic environmental settings. To numerically evalute the performance of the proposed silhouette extraction scheme, we manually segmented the human body from each video frame as the ground truth and compared the silhouette extraction results against the ground truth. Table 1 shows the average fraction of human body pixels that have been correctly included in the silhouette (True silhouette) and that have been incorrectly included into the background (False background). We can see that the proposed algorithm is very accuate.

6. CONCLUSION AND FUTURE WORK

In this work, we have described an accurate and robust human silhouette extraction scheme for activity analysis and human tracking in a dynamic environment. We calculated features from image regions and accumulate the feature information over time to build a dynamic background model. To deal with the unique challenges silhouette extraction in dynamic environments, we developed a scheme to fuse the high-level knowledge with low-level features and propose a fuzzy decision system to assure that the silhouette extraction is accurate and robust. In our future work, we will deploy the video camera system with this silhouette extraction scheme in TigerPlace, an eldercare facility at Columbia MO, and evalute its performance. We will also study the impact of silhouette extraction accuracy on the performance of activity modeling and analysis modules.

Acknowledgement

This work was supported in part by a NSF ITR grant IIS-0428420.

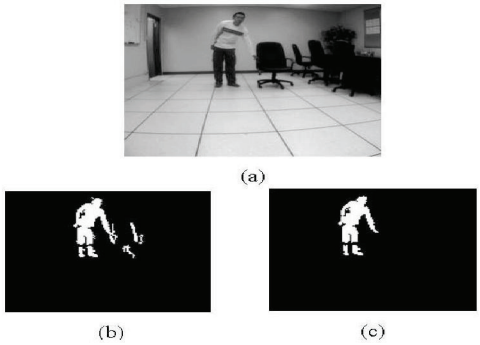

Figure 7: Detaching the moving chair from the human silhouette.

7. REFERENCES

[1] L. Wang, T. Tan, H. Ning and W. Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp.1505-1518, December 2003.

[2] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S. Nakagawa, "Automated Detection of Human for Visual Surveillance System," *Poc. Int'l Conf. Pattern Recognition,* pp. 865-869, 1996.

[3] A. Elgammal, D. Harwood, and L. Davis, "Nonparametric Model for Background Subtraction," *6th European Conference on Computer Vision,* Dublin, Ireland, July 2000.

[4] O. Schreer, I. Feldmann, U. Golz and P. Kauff, "Fast and Robust Shadow Detection in Videoconference Applications," *4th EURASIP-IEEE Region 8 international Symposium on Video/Image Processing and Multimedia Communications,* pp. 371-375, June 2002.

[5] T. Horprasert, D. Harwood, and L. Davis, "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection," *Proc. IEEE ICCV'99 FRAME-RATE Workshop,* Kerkyra, Greece, September 1999.